Stochastic stability in a learning dynamic with best response to noisy play

Christopher Kah, Markus Walzl

Working Papers in Economics and Statistics

2015-15

University of Innsbruck
Working Papers in Economics and Statistics

The series is jointly edited and published by

- Department of Banking and Finance
- Department of Economics
- Department of Public Finance
- Department of Statistics

Contact address of the editor:
Research platform “Empirical and Experimental Economics”
University of Innsbruck
Universitaetsstrasse 15
A-6020 Innsbruck
Austria
Tel: + 43 512 507 7171
Fax: + 43 512 507 2970
E-mail: eeecon@uibk.ac.at

The most recent version of all working papers can be downloaded at
http://eeecon.uibk.ac.at/wopec/

For a list of recent papers see the backpages of this paper.

Stochastic Stability in a Learning Dynamic with Best Response to Noisy Play

Christopher Kah* Markus Walzl⋆

Abstract

We propose a learning dynamic with agents using samples of past play to estimate the distribution of other players' strategy choices and best responding to this estimate. To account for noisy play, estimated distributions over other players' strategy choices have full support in the other players' strategy sets for positive levels of noise and converge to the sampled distribution in the limit of vanishing noise. Recurrent classes of the dynamic process only contain admissible strategies and can be characterised by minimal CURB sets based on best responses to noisy play whenever the set of sampled distributions is sufficiently rich. In this case, the dynamic process will always end up in a set of strategies that contains the support of a (trembling hand) perfect equilibrium. If the perfect equilibrium is unique and in pure strategies, the equilibrium resembles the unique recurrent class of the dynamic process. We apply the dynamic process to learning in matching markets and sequential two player games with perfect information.

JEL-Classification: C72; C73; D83.

Keywords: Best-response learning; equilibrium selection; stochastic stability; trembling hand perfection; CURB sets.

1 Introduction and Motivation

Learning processes have been utilised to address convergence to equilibrium and equilibrium selection in games with multiple equilibria (see, e.g., Ellison, 1993; Kandori et al., 1993; Young, 1993; Samuelson, 1994; Ellison, 2000; Peski, 2010). In these models, agents observe a limited sample of past play and choose a strategy that is optimal against the relative frequencies of other players' strategies in the

*Department of Economics, Innsbruck University.

⋆Department of Economics, Innsbruck University and METEOR, Maastricht University.

sample.[1] That is, agents are not cautious – they take the sample frequencies as the probability distribution over other players' strategy choices for the next period of the process and ignore that players may choose strategies in the future that have not been observed in the past. We propose a learning model with *best response to noisy play* that introduces an element of caution into standard learning dynamics. In this model, players not just best respond to samples of past play but to distributions that have full support in the other players' set of strategies and converge – in the limit of vanishing "noise" – to the actually observed sample. For example, players may assume that other players occasionally make a "typo" and choose an arbitrary strategy or use the observed sample to estimate a logit choice distribution of the other players' strategy choices.

We formulate a Markov process based on best responses to noisy play and analyse its recurrent classes (i.e., strategies that will be played in the long run). As long as players best respond to distributions of other players' strategies that have full support, only admissible strategies (i.e., strategies that are not (weakly) dominated) will be played in the long run. Whenever players can sample a sufficiently rich set of probability distributions over the other players strategies, the recurrent classes coincide with minimal CURB sets (see Basu and Weibull, 1991) based on best response to noisy play. In the limit of vanishing noise, we will refer to these sets as *perfectly* CURB sets.

Introducing such an anticipation of trembles or mistakes as an element of caution to a learning dynamic solves some notorious problems of best response learning. First, as already pointed out by Samuelson (1994) with the following example, players may not be able to learn the inferiority of weakly dominated strategies.

	l	r
U	1, 1	1, 0
D	1, 1	0, 0

Figure 1: A game with alternative best replies.

Example 1. In the game depicted in Figure 1, player 1 has two different best replies to l, while only U is an "always best response" (U weakly dominates D).[2] Note that starting from a state where only (U, l) is played, D is an *alternative* best response for player 1. Hence, there is no "evolutionary" pressure that prevents play moving out of the state where only admissible strategies are played. But as soon as player 1 expects player 2 to play r with a small probability ϵ, U is the

[1] In learning dynamics based on logit choice (see, e.g., Blume, 1993 or Alós-Ferrer and Netzer, 2010), agents are not assumed to best respond to sample frequencies but choose strategies according to a logit distribution over relative payoffs based on sample frequencies.

[2] We refer to the row player as *player 1* and to the column player as *player 2*.

unique best response of player 1 against any observed sample of player 2's play such that (U, l) forms the unique recurrent class of the learning dynamics.[3] $\triangle$

Second, the lack of caution in standard best response learning is also an obstacle for the learning of subgame perfect strategies in extensive form games. Whenever two players move sequentially in a generic two-stage extensive form game with perfect information (i.e., a game with no ties between any two terminal histories), the second mover has a weakly dominant strategy to pick the unique subgame perfect action at every node. This strategy becomes strictly dominant as soon as she best responds to noisy play. As a consequence, the unique subgame perfect strategy profile is a singleton recurrent class in a dynamic with best response to noisy play but not for a standard learning process.

Third, equilibria with alternative best replies (not only in weakly dominated strategies) are in general rather fragile in standard best response learning. If a player faces alternative best replies against the particular sample of past play that she observed, she plays it with a positive probability (or is likely to experiment with the alternative best response) and if other players have new best replies against this alternative best response, the corresponding equilibrium can be delearned rather easily. But this fragility seems to be at odds with data on actual play. For example, Charness and Jackson (2007) conducted a laboratory experiment where group members voted on play in a stag hunt game with varying quotas and demonstrate that players learn to coordinate on the equilibrium with alternative best replies (i.e., an equilibrium where only one out of two players has to vote for a certain strategy in order to implement it) rather than a strict equilibrium (i.e., an equilibrium where it needs the consent of both players to implement a certain strategy). This is also the unique prediction of a learning dynamic with best response to noisy play.[4] This suggests that dynamics with best response to noisy play provide a better learning framework compared to standard best response learning in games with alternative best replies (e.g., in group decision making or voting). We provide an example of a centralised one-to-many matching market where participants can coordinate on truthful preference revelation or manipulation to illustrate the relevance for market design.

Conceptually, best responses to noisy play share the idea of a robustness of strategy choices against small noise in other players' strategy choices with the concept of trembling hand perfection (see Selten, 1975). But this conceptual similarity does not yield a close relationship between trembling hand perfect equilibrium strategies and strategies in recurrent classes of the learning process for

[3] As demonstrated in Example 1, weakly dominated strategies are not necessarily eliminated by best-response learning with a fixed population. For a detailed discussion of the necessity and sufficiency of an infinite population size for the elimination of weakly dominated strategies see Kuzmics (2011).

[4] Charness and Jackson (2007) offer an equilibrium concept (*robust belief equilibrium*) that is also based on tremble anticipation to explain their findings.

arbitrary games. As it can be shown that perfectly CURB sets always contain the support of a trembling hand perfect equilibrium, the learning process will always visit trembling hand perfect equilibrium strategies in the long run for sufficiently rich sampling. However, we also show that non-equilibrium strategies can be elements of a recurrent class and trembling hand perfect equilibrium strategies can be outside any recurrent class of a learning process with best response to noisy play. Only if we restrict attention to trembling hand perfect equilibria in pure strategies, a closer relationship can be established. While a given trembling hand perfect equilibrium in pure strategies need not to be in a recurrent class of a given learning process with best response to noisy play, there is always a specification of noise such that this equilibrium resembles a singleton recurrent class of the learning process. And if the trembling hand perfect equilibrium in pure strategies is the unique perfect equilibrium of the game, it also constitutes the unique recurrent class of any learning process with best response to noisy play.

A learning dynamic with best response to noisy play introduces the concept of cautious best response (as defined in Pearce, 1984) to best response learning processes as in Young (1993). Our results also indicate a close relation between recurrent classes of a dynamic with best response to noisy play and rationalisable strategies in the presence of payoff uncertainties as discussed in Dekel and Fudenberg (1990): Only strategies that survive the procedure in Dekel and Fudenberg (1990) that first eliminates weakly dominated strategies and then (iteratively) eliminates strictly dominated strategies can be elements of recurrent classes.

Our findings contribute to a small literature on caution in learning dynamics that recently attracted renewed attention. A first approach to model caution in learning dynamics has been proposed by Hurkens (1995). In a variant of Young (1993)'s learning process where agents can sample arbitrary probability distributions over strategies in the observed sample, Hurkens (1995) establishes a one-to-one relation between recurrent classes and CURB sets. In particular, Hurkens (1995) analyses a version of this learning dynamic where agents do not best respond to the sampled probability distribution but exhibit caution by choosing a semi-robust best response (see Balkenborg, 1992), i.e., a best response against an open neighbourhood of the other players' profile. For a given learning process, a best response to noisy play is always a semi-robust best response but not vice versa. As a consequence, the recurrent classes of a dynamic with best response to noisy play are a non-empty subset of the recurrent classes in this version of Hurkens (1995)'s learning process (and thereby persistent retracts as introduced in Kalai and Samet, 1984). A complementary approach to integrate caution into decision making in games has recently been proposed by Myerson and Weibull (2015) who consider best replies against open neighbourhoods of *equilibrium* strategy profiles. The corresponding equilibria (so-called *settled equilibria*) are a selection of persistent equilibria. Another recent discussion of the relation between learning dynamics and different equilibrium notions has been offered by Balkenborg, Hofbauer, and Kuzmics (2013, 2014). They propose refined best responses (on the set of semi-robust best replies) and associated learning dynamics that make most Nash equilibria (and only the Nash equilibria) asymptotically stable. While the refined best response correspondence shares several properties such as convex-valuedness and upper hemi-continuity with a best response to noisy play, fixed points of the two correspondences may differ. In particular, no clear-cut relation to perfect equilibria can be established for dynamics based on refined best responses as a fixed point of the refined best response correspondence does not need to resemble a perfect equilibrium (see Balkenborg, Hofbauer, and Kuzmics, 2014) – in contrast to perfection of fixed points of a best response to noisy play correspondence.

The remainder of this paper is organised as follows: Section 2 introduces the learning dynamics, Section 3 discusses its recurrent classes and the relation to CURB sets. Section 4 focuses on the relation between recurrent classes and perfect equilibria and Section 5 extends the analysis to stochastic stability. In Section 6 we conclude with some remarks on applications and related solution concepts. Proofs and some auxiliary results are relegated to Appendix A.

2 Set-up

Let $\Gamma = (N, (S_i)_{i\in N}, (u_i)_{i\in N})$ denote an N-player strategic form game with a finite set of players N, a finite set of pure strategies for player $i \in N$, S_i, and payoffs u_i. Let $S = \prod_{j\in N} S_j$ be the Cartesian product of players' pure strategy sets and let $\Theta_i := \Delta(S_i)$ be the set of probability distributions over S_i such that the mixed extension of Γ is given by $(N, (\Theta_i)_{i\in N}, (U_i)_{i\in N})$ with player i's von Neumann-Morgenstern utility U_i. Further, denote $\Theta := \prod_{i\in N} \Theta_i$ and let $\Theta_{-M} := \prod_{j\in N\setminus M} \Theta_j$, with $M \subsetneq N$. $\sigma_i(s_i)$ denotes the probability mass that some mixed strategy $\sigma_i \in \Theta_i$ assigns to pure strategy $s_i \in S_i$ of player $i \in N$. We call a strategy $\sigma_i \in \operatorname{int}(\Theta_i)$ (i.e., $\sigma_i(s_i) > 0$ for each $s_i \in S_i$) *fully mixed*.[5] A *Nash equilibrium* is a strategy profile $\sigma^* \in \Theta$ such that for all $i \in N$ and $\sigma_i \in \Theta_i \setminus \{\sigma_i^*\}$: $U_i(\sigma_i^*, \sigma_{-i}^*) \geq U_i(\sigma_i, \sigma_{-i}^*)$. A Nash equilibrium $\sigma^* \in \Theta$ is *strict* if for all $i \in N$ the former inequality holds strictly. We denote the *best response correspondence* by $\beta := \prod_{i\in N} \beta_i$ with $\beta_i : \Theta_{-i} \rightrightarrows \Theta_i$ for each $i \in N$, i.e., $\sigma_i' \in \beta_i(\sigma_{-i})$ implies $\sigma_i' \in \arg\max_{\sigma_i\in\Theta_i} U_i(\sigma_i, \sigma_{-i})$. Further, let $\mathcal{B}_i(\sigma_{-i}) := \bigcup_{\sigma_i\in\beta(\sigma_{-i})} \operatorname{supp} \sigma_i$ denote the set of player i's pure strategy best replies against σ_{-i} and set $\mathcal{B}(\sigma) := \prod_{i\in N} \mathcal{B}_i(\sigma_{-i})$. A (possibly pure) strategy $\sigma_i \in \Theta_i$ is *strictly dominated* whenever there is $\sigma_i' \in \Theta_i$ such that $U_i(\sigma_i', \sigma_{-i}) > U_i(\sigma_i, \sigma_{-i})$ for all $\sigma_{-i} \in \Theta_{-i}$. σ_i is *undominated* whenever it is not strictly dominated. A (possibly pure) strategy $\sigma_i \in \Theta_i$ is *weakly dominated*

[5] Strictly speaking, we refer to the *relative interior* of a convex set (relative to its affine hull).

whenever the preceding inequality holds weakly for each $\sigma_{-i} \in \Theta_{-i}$, and strictly for at least one such σ_{-i}. σ_i is *admissible* whenever it is not weakly dominated.

2.1 Sampling and estimation

We consider a learning process where, in each period, agents taken from a large, finite set assume the role of each player $i \in N$, observe a sample of past play, estimate the distribution of strategy choices by other players from this sample, and best respond to the estimated distribution. Formally, agents observe a sample of $k < m$ observations of play as occurred in the most recent m periods. Let $h := (s(t-m), \ldots, s(t-1))$ denote the *history* of past play (at date t), and for each $j \in N$, let $\pi_j(h) := \{s_j(t-m), \ldots, s_j(t-1)\}$ be the projection of h on S_j. Hence, $\pi_j(h)$ lists all pure strategy choices made by player j in history h. A *sampled distribution* of player i regarding the strategy choice of player j after history h is a mapping $q_{ij}(h) : S_j \to \Delta(S_j)$ that assigns to each strategy $s_j \in S_j$ its observed relative frequency in the sample drawn from history h; specifically, $q_{ij}(h)(s_j) = 0$ for all $s_j \in S_j \setminus \{\pi_j(h)\}$. For example, for $m = 3$ and $k = 2$ in the game in Figure 1, the sample (l, l) out of history (l, l, r) that player 1 observes regarding choices of player 2 induces a sampled distribution that assigns probability 1 to l and probability 0 to r. Based on the sampled distribution q_{ij}, agent i forms an *estimate* $\psi_{ij} : \Delta(S_j) \to \Delta(S_j)$ regarding the distribution of player j's strategy choices.[6] Denote player i's estimates by $\psi_i := \prod_{j \neq i} \psi_{ij}$. We will refer to $\psi = (\psi_i)_{i=1,\ldots,N}$ as an *estimation procedure*.

Procedure 1. *Learning as in Young (1993):* In the learning framework as introduced in Young (1993), each agent i considers the sampled probability distribution over past play as the actual (estimated) probability distribution over play in the relevant period (and best responds to this distribution), i.e., $\psi_{ij}(q_{ij}) = q_{ij}$ for all i, j.

In the sequel, we will be interested in estimates $\psi_{ij}(q_{ij}, \epsilon)$ that are noisy in the sense that the estimated probability distribution of player i regarding the choice of player j has full support in S_j for positive levels of noise $\epsilon > 0$ and converges against the sampled distribution in the limit of vanishing noise $\epsilon \to 0$.

Definition 1. An estimation procedure ψ is called **noisy**, if for all $i, j \in N$ and $\sigma_j \in \Delta(S_j)$ (i) $\operatorname{supp}\left(\psi_{ij}(\sigma_j, \epsilon)\right) = S_j$ for $\epsilon > 0$ and $\psi_{ij}(\sigma_j, \epsilon) = \sigma_j$ for $\epsilon = 0$, and (ii) $\psi_{ij}(\sigma_j, \epsilon)$ is jointly continuous in σ_j and ϵ.

The following two estimation procedures are noisy.

[6] As ψ_{ij} maps sampled distributions into estimated distributions, its pre-image for a given history h is $\Delta(\pi_j(h))$. But as we will address the continuity of the estimation process in $\Delta(S_j)$ and derive results for rich sampling – i.e., when an arbitrary probability distribution can be a sampled distribution – we formally define ψ_{ij} on $\Delta(S_j)$.

Procedure 2. *Anticipation of typos:* Suppose player i samples the distribution $\sigma_j \in \Delta(S_j)$ and expects player j to choose pure strategy s_j according to σ_j with probability $(1-\epsilon)$ and to play according to a uniformly mixed strategy τ_j with probability $\epsilon > 0$. Observe that the linearity in ϵ guarantees continuity in ϵ and σ_j.

Procedure 3. *Logit choice:* Suppose player i samples the distribution $\sigma_j \in \Delta(S_j)$ and expects player j to choose a strategy s_j according to σ_j with probability $(1-\epsilon)$ and to tremble according to a logit distribution τ_j with probability $\epsilon > 0$. I.e., if player j trembles, $s_j \in S_j$ is chosen with probability $\tau_j(s_j) = \frac{e^{\lambda u_j(s_j,\sigma_{-j})}}{\sum_{s\in S_j} e^{\lambda u_j(s,\sigma_{-j})}}$ where σ_{-j} is the sampled distribution of play for all players $k \neq j$ as observed by player i. λ is a standard logit parameter. Observe that τ_j has full support in S_j for any finite λ, and that $(1-\epsilon)\sigma_j + \epsilon\tau_j$ is continuous in ϵ and σ_j.

2.2 Best response to noisy play

A strategy $s_i \in S_i$ is a *best response to noisy play* against $\sigma_{-i} \in S_{-i}$ if s_i is a best response to the estimated distribution $\psi(\sigma_{-i}, \epsilon)$.

Definition 2. For $i \in N$ and $\sigma_{-i} \in \Theta_{-i}$, $\beta_i^\epsilon : \Theta_{-i} \rightrightarrows \Theta_i$ with $\beta^\epsilon(\sigma_{-i}) = \beta(\psi(\sigma_{-i}, \epsilon))$ is a **best response to noisy play correspondence**.

As $\psi(\sigma_{-i}, \epsilon)$ is continuous in $\sigma_{-i} \in \Theta_{-i}$ and the best response correspondence β is upper hemi-continuous (u.h.c.), non-empty, and convex-valued, the best response to noisy play correspondence is also u.h.c., non-empty, and convex-valued (see, e.g., Aliprantis and Border, 2006, Theorem 17.23).

Lemma 1. *A best response to noisy play correspondence is u.h.c., non-empty, and convex valued.*

In Procedure 2, for instance, any set $T_{-i} \in \mathcal{P}(N \setminus \{i\})$ is a possible set of trembling players according to player i's expectations.[7] The probability that a given set T_{-i} of players tremble when each player trembles with probability ϵ therefore reads

$$\mathbb{P}(T_{-i}, \epsilon) = \epsilon^{|T_{-i}|} (1-\epsilon)^{|N\setminus\{i\}|-|T_{-i}|}. \tag{2.1}$$

Now let $\tau(T_{-i}) \in \operatorname{int}\left(\prod_{j\in T_{-i}} \Theta_j\right)$ be uniformly mixed for every player $j \neq i$. Then, a strategy σ_i is a best response to noisy play, i.e., $\sigma_i \in \beta_i^\epsilon(\sigma_{-i})$, if σ_i solves

$$\max_{\sigma_i'\in\Theta_i} \left(1 - \sum_{T_{-i}} \mathbb{P}(T_{-i}, \epsilon)\right) U_i(\sigma_i', \sigma_{-i}) + \sum_{T_{-i}} \mathbb{P}(T_{-i}, \epsilon) U_i\left(\sigma_i', \sigma_{-\{\{i\}\cup T_{-i}\}}, \tau(T_{-i})\right). \tag{2.2}$$

For a given $\epsilon \geq 0$, the linearity of U_i in σ_i' enures that the best response to noisy play correspondence is u.h.c., non-empty, and convex valued.

[7] Let $|X|$ denote the cardinality of some set X and let $\mathcal{P}(X) := 2^X \setminus \{\varnothing\}$.

2.3 Markov process

For a given noise level ϵ, a memory length m, and sample length k, playing best response to noisy play establishes a finite time-homogeneous Markov chain $M^\epsilon := M\left(\Omega, [\mathbb{P}(\omega,\omega')]_{\omega,\omega'\in\Omega}\right)$. The state space Ω is the set of histories (of length m), i.e. a generic state $\omega(t) \in \Omega$ at date t is a list $\omega \equiv (s(t-m),\ldots,s(t-1))$. The transition matrix $[\mathbb{P}(\omega,\omega')]_{\omega,\omega'\in\Omega}$ gives the transition probabilities $\mathbb{P}(\omega,\omega')$ between any two (not necessarily distinct) $\omega, \omega' \in \Omega$. A state ω' is called a *successor* of a state ω whenever $\mathbb{P}(\omega,\omega') > 0$. ω' is thereby obtained from ω by deleting the leftmost element and adding a new rightmost element. For future reference, denote the leftmost element of some $\omega \in \Omega$ as $l(\omega)$ and the rightmost element of ω as $r(\omega)$. In what follows, we will analyse this Markov process in a standard way and start with a characterisation of its recurrent classes.

3 Recurrent classes

For a dynamic process M, a (possibly singleton) set of states $A \subseteq \Omega$ is a *recurrent class* (or absorbing) set whenever A is a minimal non-empty set of states that once entered is never left by the process. If $\omega \in A$ for some absorbing set A, then ω is called *recurrent*, otherwise it is called *transient*. We call a strategy profile $s \in S$ recurrent whenever there exists a recurrent class A such that $s \in \bigcup_{\omega\in A} r(\omega)$. For further reference, let us denote the collection of recurrent classes of process M^ϵ by $\mathcal{A}^\epsilon$.

3.1 Admissibility

If agents best respond to a probability distribution over the other players' strategies, only undominated (or rationalisable) strategies can be elements of a recurrent class. However, as already discussed in Samuelson (1994), a weakly dominated strategy can well be an element of a recurrent class if estimated distributions of play do not have full support (see, e.g., Procedure 1 applied to Example 1). But as soon as the estimated distribution of play has full support (as for noisy estimation procedures), a recurrent strategy has to be admissible.

Proposition 1. *Consider a dynamic with best response to noisy play M^ϵ with a set of recurrent classes $\mathcal{A}^\epsilon$. Then, for all $i \in N$ and $\omega \in \mathcal{A}^\epsilon$, $s_i \in \pi_i(\omega)$ is admissible.*

Proof. Observe that $s_i \in S_i$ with $s_i \in \beta_i(\sigma_{-i})$ for some $\sigma_{-i} \in \Theta_{-i}$ implies that s_i is admissible whenever σ_{-i} has full support in S_{-i}. Hence, in M^ϵ there is only a transition from ω into ω' if ω' has admissible strategies as its rightmost elements and non-admissible strategies can never be recurrent. ■

By Proposition 1, admissibility of a strategy is a necessary condition for being recurrent. But admissibility or being in the support of a Nash equilibrium in admissible strategies is not sufficient to be an element of a recurrent class.

First, as long as the sampling is incomplete in the sense that only a restricted set of probability distributions over other players' play can be observed, recurrent classes may consist of non-equilibrium strategies (even vis-à-vis a unique strict Nash equilibrium in pure strategies) as demonstrated in the following example.

	l	m_1	m_2	r
U	$0,0$	$0,1$	$0,0$	$15,15$
M	$0,10$	$0,10$	$10,0$	$0,8$
D	$1,0$	$10,0$	$0,10$	$0,8$

Figure 2: Non-equilibrium strategies in a recurrent class due to limited sampling.

Example 2. Consider the game Γ_2 depicted in Figure 2 and a best response to noisy play dynamic induced by Procedure 2 with $m = 2$ and $k = 1$. For ϵ sufficiently small, the state $\omega^* = ((U,r),(U,r))$ where the strict and unique Nash equilibrium (U,r) is played for ever is recurrent. But as $m_1 = \beta_2^\epsilon(M)$, $m_2 = \beta_2^\epsilon(D)$, $M = \beta_1^\epsilon(m_2)$, and $D = \beta_1^\epsilon(m_1)$ also a cycle with m_1, m_2, M, and D is recurrent. △

If the sampling is sufficiently rich in the sense that a sufficiently large set of probability distributions can be sampled, non-equilibrium cycles as in Figure 2 cease to be recurrent. For instance, the cycle in Figure 2 is only recurrent if player 2 has a sample of size one that either displays M (with unique best response to noisy play m_1) or D (with unique best response m_2). As soon as the sample (M,D) can be observed, playing r becomes superior.

But also the richness of the set of sampled distributions can be a reason for non-equilibrium strategies to be part of a recurrent class as discussed in the following example.

	l	m	r
U	$3,-5$	$1,2$	$-1,4$
M	$0,-5$	$1,0$	$0,1$
D	$-5,-5$	$0,1$	$1,0$

Figure 3: Non-equilibrium strategies in a recurrent class due to alternative best replies.

Example 3. Consider the game Γ_3 depicted in Figure 3 and a best response to noisy play dynamic as in Procedure 2. In Γ_3 there is a unique Nash equilibrium σ^* with $\sigma_1^* = \left(\frac{1}{2}M + \frac{1}{2}D\right)$ and $\sigma_2^* = \left(\frac{1}{2}m + \frac{1}{2}r\right)$. But the non-equilibrium strategy profiles (U,m) and (U,r) are played with positive probability in the only absorbing set: When player 1 sufficiently often samples the play of m by player 2, her optimal choice is U (recall that Procedure 2 considers uniform trembles), while player 2 in turn will choose r if she samples sufficiently many observations of U. △

The fact that non-equilibrium strategies can be best responses against specific samples of equilibrium strategies has already been acknowledged by Basu and Weibull (1991) and led them to the introduction of CURB sets, i.e., sets of strategies that are closed under rational behaviour. These CURB sets have been utilised by Hurkens (1995) to characterise absorbing sets of a standard best response dynamic as in Procedure 1 where players best respond to arbitrary probability distributions over the sampled strategies. In the following section, we will adapt this characterisation to a learning process with best response to noisy play.

3.2 Perfectly CURB sets

For each $i \in N$ and $Y_i \subseteq S_i$ (with $Y := \prod_{i \in N} Y_i$), let $\Upsilon_i := \Delta(Y_i)$. Further, we denote the image of β_i^ϵ restricted to Υ_{-i} by $\beta_i^\epsilon(\Upsilon_{-i}) := \bigcup_{\sigma_{-i} \in \Upsilon_{-i}} \beta_i^\epsilon(\sigma_{-i})$ and let $\beta^\epsilon(\Upsilon) := \prod_{i \in N} \beta_i^\epsilon(\Upsilon_{-i})$.

Definition 3. A set of pure strategy profiles $Y \subseteq S$ is **β^ϵ-CURB** if $\beta^\epsilon(\Upsilon) \subseteq \Upsilon$. It is a **minimal β^ϵ-CURB set** if there does not exist some $Y' \subsetneq Y$ such that Y' satisfies $\beta^\epsilon(\Upsilon') \subseteq \Upsilon'$. For a minimal β^ϵ-CURB set $Y \subseteq S$, a strategy profile $y \in Y$ is called a **β^ϵ CURB strategy profile.** If $Y \subseteq S$ is β^ϵ-CURB for $\epsilon \to 0$, it is called a **perfectly CURB set** and strategy profiles are referred to as **perfectly CURB strategy profiles.**

According to Definition 3, a (minimal) β^ϵ-CURB set $Y \subseteq S$ is a (minimal) set of pure strategy profiles that contains, for all $i \in N$, the support of *all* the best replies according to β_i^ϵ to *any* probability distribution over Y_{-i}, i.e., $\mathcal{B}^\epsilon(\Upsilon) \subseteq Y$. Hence, the dynamic process will never return to strategies outside Y once the history only contains strategy profiles in Y. For further reference, we summarise this observation in the following lemma.

Lemma 2. *Let M^ϵ be a dynamic process with best response to noisy play. Then, any β^ϵ-CURB set $Y \subseteq S$ owns a recurrent class of M^ϵ.*

But a recurrent class does not necessarily overlap with any minimal β^ϵ-CURB set as long as agents can only estimate a limited set of probability distributions over strategies of other players. This is illustrated by Example 2 where the cycle with player 1 playing M and D and player 2 playing m_1 and m_2 was recurrent for $m = 2$ and $k = 1$. As r is the unique best response to the probability distribution that assigns m_1 and m_2 probability one-half, the cycle strategies cannot constitute a minimal β^ϵ-CURB set. In fact, it is easy to check that only the unique Nash equilibrium (U, r) resembles a β^ϵ-CURB set in the example. Hence, for $k = 1$ and $m = 2$, we have a recurrent class that does not overlap with any minimal β^ϵ-CURB set. In contrast, only the unique minimal β^ϵ-CURB set based on (U, r) is recurrent for $k = 2$ and $m = 3$. In the next section, we will discuss in how far a sufficiently rich set of sampled distributions always allows for a characterization of recurrent classes by minimal β^ϵ-CURB sets.

3.3 Rich sampling

For Procedure 1, Hurkens (1995) demonstrates that CURB sets and recurrent classes of a learning process coincide if players estimate *any* probability distribution over sampled strategies with a positive probability. But as already pointed out in Young (1998, p. 167), Hurkens (1995)'s assumption that *any* probability distribution will be estimated with a positive probability is somewhat indispensable if one wants to establish the result for all games (see also the discussion in Hurkens, 1995, p. 311). Specifically, Young (1998) provides an example where CURB sets and recurrent classes only coincide if agents estimate a probability distribution over other players' strategies that (partially) assigns irrational probability masses – which is impossible in Procedure 1, as the relative frequencies of strategies in the sample are rational by construction. Young (1998) discusses a generic class of games that are *non-degenerate in best replies* for which CURB sets and recurrent classes coincide if the sample size and the sample to memory ratio is such that a sufficiently rich set of probability distributions can be sampled. A game is non-degenerate in best replies if $\beta_i^{-1}(s_i)$ is either empty or open in the relative topology of $\prod_{j\neq i}\Theta_j$. Intuitively, whenever the pre-image of β_i for $s_i \in S_i$ is empty or open, s_i is a best response to a distribution over S_{-i} that can be sampled. To provide conditions for the coincidence of β^ϵ-CURB sets and recurrent classes of a dynamic with best response to noisy play, we adapt the definition of games that are non-degenerate in best replies to the following weaker requirement.

Definition 4. A game is **weakly non-degenerate in best replies** if for every i and $s_i \in S_i$, $(\beta_i^\epsilon)^{-1}(s_i)$ is either empty or the projection of $(\beta_i^\epsilon)^{-1}(s_i)$ to Θ_j is, for all $j \neq i$, either open in Θ_j or a pure strategy.

In contrast to the definition in Young (1998, p. 165), we do not exclude games where a certain strategy of player i is only a best response against a certain pure strategy profile of other players, e.g., a situation where player i has a weakly dominated strategy.

Strategies that are best replies to an open subset of strategy profiles (i.e., robust best replies as defined in Balkenborg, 1992) have also been discussed in Balkenborg, Hofbauer, and Kuzmics (2013). The following definition is taken from Kalai and Samet (1984).

Definition 5. Two strategies σ_i and σ_i' are *own-payoff equivalent* if $U_i(\sigma_i, \sigma_{-i}) = U_i(\sigma_i', \sigma_{-i})$ for all $\sigma_{-i} \in \Theta_{-i}$.

Balkenborg, Hofbauer, and Kuzmics (2013) demonstrate that in games without own-payoff equivalent strategies, a strategy is a best reply to an open subset of strategy profiles if and only if it is not weakly inferior. A weakly inferior strategy is never the only best reply and in two-player games a strategy is weakly inferior if and only if the strategy is either weakly dominated of equivalent to a proper mixture of pure strategies.

Whenever a game is weakly non-degenerate in best replies and the sample and memory size allows to elicit a sufficiently rich set of probability distributions over strategies, recurrent classes of a dynamic process with best response to noisy play and the corresponding β^ϵ-CURB sets coincide.

Lemma 3. *Consider a game that is weakly non-degenerate in best replies and let M^ϵ be a dynamic process with best response to noisy play. Then, there are $\widetilde{k}$ and $\widetilde{m}$ such that recurrent strategies and minimal β^ϵ-CURB strategies coincide for any $k \geq \widetilde{k}$ and $m \geq \widetilde{m}$.*

Proof. See Appendix A. ∎

Definition 6. A dynamic process M^ϵ with sample and memory sizes that satisfy $k \geq \widetilde{k}$ and $m \geq \widetilde{m}$ as in Lemma 3 is referred to as a dynamic process with **rich sampling**.

4 Relation to trembling hand perfection

4.1 Trembling hand perfection

Conceptually, a noisy estimation procedure is closely related to (trembling hand) perfection, i.e., the robustness of equilibrium strategies against small (anticipated) trembles.

Definition 7. In a game Γ, $\sigma^p \in \Theta$ is a **perfect equilibrium** if and only if there is some sequence $\left\{\sigma^\ell\right\}_{\ell=1}^{\infty} \subseteq \operatorname{int}(\Theta)$, with $\left\{\sigma^\ell\right\}_{\ell=1}^{\infty} \to \sigma^p$, such that for each ℓ and each $i \in N$, $\sigma_i^p \in \beta_i\left(\sigma_{-i}^\ell\right)$.

The following game taken from van Damme (1991) illustrates how (trembling hand) perfection and a dynamic process with best response to noisy play select strategy profiles based on the same reasoning.

M_1:

	l	r
U	$1,1,1$	$1,0,1$
D	$1,1,1$	$0,0,1$

M_2:

	l	r
U	$1,1,0$	$0,0,0$
D	$0,1,0$	$1,0,0$

Figure 4: A 3-player game with two equilibria in admissible strategies. Player 1 chooses, rows, player 2 chooses columns, and player 3 chooses matrices.

Example 4. Consider the 3-player game Γ_4 depicted in Figure 4.[8] Both $s^* = (U, l, M_1)$ and $s' = (D, l, M_1)$ are pure strategy Nash equilibria in admissible strategies but only s^* is (trembling hand) perfect as U is a best response against fully mixed strategies of players 2 and 3 that converge to l and M_1, respectively, while D is not. Likewise the state $\omega' = (s', \ldots, s')$ will be transient and $\omega^* = (s^*, \ldots, s^*)$ will be the unique absorbing state in a dynamic process of best response to noisy play as in Procedure 2. To the contrary, play according to conventional best response as depicted in Procedure 1 does not converge to a singleton absorbing set as starting from ω^* there is a positive probability that the process moves on to ω'' with $r(\omega'') = s'$. △

4.2 Recurrent classes and perfect pure strategy equilibria

By construction of the learning dynamic, a strategy profile that constitutes a singleton recurrent class of a dynamic process with best response to noisy play in the limit of vanishing noise is also a trembling hand perfect equilibrium in pure strategies. But looking at Definition 7 also reveals that the selection of trembling hand perfect equilibria by the dynamic process in Example 4 will hardly generalise to arbitrary games – just note that Selten's definition puts *no* restriction on the relative trembling probabilities, while the dynamic process works with a fixed specification. As a consequence, the support of a trembling hand perfect equilibrium (even in pure strategies) may not be in any recurrent class as the following example demonstrates.

	l	m_1	m_2	r
U	5, 2	3, 0	0, 3	6, 4
D	5, 4	0, 3	2, −1	6, −10

Figure 5: A game with two trembling hand perfect equilibria in pure strategies. Only one equilibrium constitutes the unique singleton recurrent class of the dynamics M^ϵ.

Example 5. For the game Γ_5 depicted in Figure 5 the pure strategy profiles (D, l) and (U, r) are perfect Nash equilibria. To illustrate, consider Procedure 2 and fix $m = 2$ and $k = 1$. Then, U is the unique best response to noisy play of player 1 to all samples except $\{m_2\}$ and the only recurrent state is

$$\omega^p = \begin{pmatrix} U & U \\ r & r \end{pmatrix}$$

△

But for the example it is also easy to find a dynamic process with best response to noisy play such that the other trembling hand perfect equilibrium in

[8] Van Damme (1991) uses this example to show that trembling hand perfection might reject admissible strategies in games with more than two players.

pure strategies constitutes the unique singleton recurrent class of the process. Just consider a slightly modified specification of Procedure 2 where with probability ϵ player 1 expects player 2 to play a fully mixed strategy where r and m_2 are played with probability $1/3$ each and m_1 and l with probability $1/6$ each. Then, m_2 is twice as likely as m_1 for a given $\epsilon > 0$ and D is the unique best response to noisy play of player 1 except for the sample $\{m_1\}$. As a consequence, (D, l) is the unique (singleton) recurrent class in this case.

As the following result indicates, this observation generalises to arbitrary games as long as no player has own-payoff equivalent strategies.

Proposition 2. *Let Γ be a game with a perfect equilibrium s^p in pure strategies, and suppose there is no player i with a strategy s_i' that is own-payoff equivalent to s_i^p. Then, there exists a dynamic process with best response to noisy play M^ϵ such that $\omega = (s^p, \ldots, s^p)$ constitutes a singleton recurrent class whenever $\epsilon > 0$ is sufficiently small.*

Proof. See Appendix A. ■

Whenever there is a player i with an own-payoff equivalent strategy to s_i^p, it is impossible to find a process such that s^p constitutes a singleton recurrent class (as no expectations about opponents' play ever prevent that an equivalent strategy is a best response (to noisy play) whenever the original strategy is). Moreover, as long as there is another player whose best response depends on which of the equivalent strategies is used by player i, s^p may become transient. But as long as we exclude strategies that are own-payoff equivalent to s^p for any player, we can always translate the sequence of fully mixed strategies that s^p best responds to (see Definition 7) into a dynamic process where s^p is a best response to noisy play. If potential alternative best replies are not own-payoff equivalent strategies, we can always augment the dynamic process with additional small noise such that s^p is the unique best response to noisy play against itself. As a consequence, s^p constitutes a singleton recurrent class.

For trembling hand perfect equilibria in mixed strategies, however, no such general relation between the support of an equilibrium and recurrent strategies can be established. First, the support of a trembling hand perfect mixed strategy equilibrium may not establish a minimal (perfectly) CURB set (and thereby a recurrent class for sufficiently rich sampling). This is straightforward to see in a game with a trembling hand perfect equilibrium in mixed strategies that contains a strict Nash equilibrium in its support (consider, for instance, the Battle-of-Sexes). Such a mixed strategy equilibrium will never be minimally (perfectly) CURB (and thereby fail to resemble a recurrent class for sufficiently rich sampling as indicated by Lemma 3). Second, a minimally (perfectly) CURB set may exceed the support of a trembling hand perfect equilibrium as illustrated by Example 3. Here, the unique minimal (perfectly) CURB set Y is given by $Y = \{\{U, M, D\}, \{m, r\}\}$ and contains non-equilibrium strategies.

4.3 Path to trembling hand perfection

As discussed in the previous section, a strategy being in the support of a trembling hand perfect equilibrium (in pure strategies or not) is neither necessary nor sufficient for this strategy to be in a recurrent class – there may be off-equilibrium strategies that are recurrent and trembling hand perfect equilibrium strategies that fail to be recurrent. But we can build on the facts that CURB sets characterise recurrent classes (for sufficiently rich sampling) and that CURB sets contain the support of a Nash equilibrium (see Basu and Weibull, 1991) to demonstrate that minimally perfectly CURB sets (and thereby – for rich sampling – recurrent classes) contain the support of a trembling hand perfect equilibrium.

Proposition 3. *Let Γ be a game and $Y \subseteq S$ a minimally perfectly CURB set. Then, there exists a perfect equilibrium σ^p of Γ with* $\operatorname{supp}(\sigma^p) \subseteq Y$.

Proof. See Appendix A. ■

The idea of the proof is first to show that fixed points of β^ϵ induce perfect equilibria as $\epsilon \to 0$. We then consider a convergent sequence of fixed points of β^ϵ with supports in β^ϵ-CURB sets and show that the limit is a perfect equilibrium, i.e., a perfectly CURB set owns the support of a perfect equilibrium. As a consequence, a dynamic process with best response to noisy play (that only visits perfectly CURB strategies in the long run and for small noise) will visit strategies in the support of a trembling hand perfect equilibrium – not necessarily exclusively but with a strictly positive probability. In this sense, we get a *path to trembling hand perfection.*

Corollary 1. *Let Γ be a game that is weakly non-degenerate in best replies and M^ϵ be a dynamic process with best response to noisy play and rich sampling. Then, any recurrent class of M^ϵ owns the support of a perfect equilibrium if $\epsilon > 0$ is sufficiently small.*

4.4 Unique perfect equilibria in pure strategies

As perfect equilibria in pure strategies may not constitute recurrent classes in the presence of other perfect equilibria (see Example 5) and CURB sets may exceed the support of unique perfect mixed strategy equilibria (see Example 3), we are left with games that exhibit a unique perfect equilibrium in pure strategies. In this case, we can establish the following one-to-one relation between recurrent classes and perfect equilibria.

Proposition 4. *Let Γ be a game with a unique perfect equilibrium $s^p \in S$ that is in pure strategies. Then, for $\epsilon > 0$ sufficiently small,*

(i) s^p constitutes a singleton recurrent class $\omega^p = (s^p, \ldots, s^p)$ of a learning dynamic with best response to noisy play M^ϵ;

(ii) if Γ is weakly non-degenerate in best replies and M^ϵ a process with rich sampling, $\mathcal{A}^\epsilon = \{\omega^p\}$, i.e., s^p is the unique recurrent strategy profile.

Proof. See Appendix A. ∎

The proof is based on the fact that a unique perfect equilibrium is *strictly perfect.*[9] Strict perfection together with the uniqueness of s^p then implies that $\mathcal{B}^\epsilon(s^p)$ is singleton-valued – even if there are own-payoff equivalent strategies. Were $\mathcal{B}^\epsilon(s^p)$ not singleton-valued, we could construct additional perfect equilibria. But if s^p constitutes a singleton recurrent class, it is also a a minimal (perfectly) CURB set whenever the sampling is rich. But as (by Proposition 3) any perfectly CURB set contains the support of a trembling hand perfect equilibrium, equilibrium uniqueness then also implies that the recurrent class with s^p played for ever is the only recurrent class of the process.

Proposition 4 is related to Hurkens (1995, Corollary 2) where it is shown that a conventional best response dynamic converges to the *unique strict* equilibrium of a game in the case of rich sampling. Anticipating noise in the process allows play to converge to a unique perfect equilibrium in pure strategies that is not necessarily strict.

5 Stochastic stability

In learning dynamics with multiple recurrent classes it is customary to analyse the stochastic stability of these recurrent classes, i.e., the recurrent classes vulnerability to small mistakes or exogenous shocks. For example, in Young (1993) a player randomly chooses a strategy with probability ϵ in any given period and a recurrent class is stochastically stable if it is at least as robust against these random choices as any other recurrent class. A straightforward way to implement shocks or mistakes in a dynamic process with best response to noisy play is to assume that anticipated noise actually occurs. Consider, e.g., Procedure 2 and assume that players indeed commit the anticipated typos. The resulting Markov chain $\widetilde{M}^\epsilon$ is irreducible and aperiodic for any $\epsilon > 0$ and therefore exhibits a unique invariant distribution. An invariant distribution $\mu \in \Delta(\Omega)$ is a probability distribution over states that is not altered by the processes' transition matrix (i.e. a

[9] A strictly perfect equilibrium consists of best replies against *arbitrary* sequences of small trembles. For a formal treatment, see Okada (1981) or Definition A4 in the Appendix.

vector of dimension $|\Omega|$ such that $\mu = \mu \cdot [\mathbb{P}^{\epsilon}(\omega, \omega')]_{\omega,\omega' \in \Omega}$). For irreducible and aperiodic Markov chains the component of μ corresponding to some $\omega \in \Omega$ gives the probability to be in ω as $t \to \infty$.

We are now interested in the case $\epsilon \to 0$, i.e., the limit of small perceived *and* actual noise. The corresponding invariant distribution is the *limit* invariant distribution $\mu^* = \lim_{\epsilon \to 0} \mu$ and its support is the set of *stochastically stable states*.

Definition 8. Let $\widetilde{M}^{\epsilon}$ be an irreducible and aperiodic Markov chain.

(i) $\omega \in \Omega$ is *stochastically stable* whenever $\mu^*(\omega) > 0$. Define $\Omega^* := \{\omega \in \Omega : \omega \in \mathrm{supp}(\mu^*)\}$.

(ii) A pure strategy $s_i \in S_i$ is stochastically stable whenever there is a state $\omega \in \Omega^*$ with $s_i \in \pi_i(\omega)$.

Typically, the identification of the set Ω^* relies on graph-theoretic techniques as put forward by Freidlin and Wentzell (1984). The key insight is that $\omega \in \Omega^*$ if and only if $\omega \in \arg\min_{\omega' \in \Omega} \gamma(\omega')$, where $\gamma(\cdot)$ denotes the *stochastic potential.* The *stochastic potential* of a state ω is the minimum number of mistakes it takes to construct a spanning tree in Ω rooted in ω with edges between two states $\omega', \omega'' \in \Omega$ only if $\mathbb{P}^{\epsilon}(\omega', \omega'') > 0$. Intuitively, the stochastic potential of a state ω is the smaller the easier it is to reach ω from all other states via process M^{ϵ} and the more difficult it is to exit ω. [10] As already recognised in Theorem 4 of Young (1993), only absorbing sets of the "unperturbed" process M^{ϵ} are candidates for stochastically stable states and absorbing states of M^{ϵ} are stochastically stable states if and only if they have minimal stochastic potential.

Lemma 4. *Let $\mathcal{A}^{\epsilon}$ be the collection of absorbing sets of a dynamic process with best response to noisy play M^{ϵ}. Then:*

(i) $\Omega^* \subseteq \mathcal{A}^{\epsilon}$.

(ii) $\omega \in \Omega^*$ *if and only if* ω *has minimal stochastic potential in* $\widetilde{M}^{\epsilon}$.

A direct implication of Lemma 4 is that (i) stochastically stable strategies are admissible (see Proposition 1), (i) the support of the limit invariant distribution contains the support of a perfect equilibrium (under the conditions of Proposition 3), and a unique perfect equilibrium in pure strategies is the unique stochastically stable strategy profile (under the conditions of Proposition 4).

Note, however, that stochastic stability for a dynamic with best response to noisy play does not "refine" stochastic stability for conventional best response learning (as in Procedure 1). To see this consider the following example.

[10] See Vega-Redondo (2003, Ch. 12) for a textbook treatment of these techniques.

	l	m	r
U	$10,10$	$2,2$	$1,1$
M	$10,1$	$0,2$	$0,1$
D	$0,0$	$3,0$	$4,4$

Figure 6: A game with different stochastically stable strategies for M^ϵ and conventional best response learning.

Example 6. Consider the game Γ_6 depicted in Figure 6. The strategy profiles (U,l) and (D,r) are perfect equilibria in pure strategies. Consider first $m = 2$ and $k = 1$ with states

$$\omega^p = \begin{pmatrix} U & U \\ l & l \end{pmatrix} \quad \text{and} \quad \omega^{p'} = \begin{pmatrix} D & D \\ r & r \end{pmatrix}.$$

Under conventional best response learning we get $\left\{\omega^{p'}\right\}$ as the unique recurrent class because M is an alternative best reply to l. But for a best response to noisy play dynamic M^ϵ, U becomes a unique best response such that the set of recurrent classes is $\left\{\omega^p, \omega^{p'}\right\}$. By Lemma 4, $\left\{\omega^{p'}\right\}$ is then also the unique stochastically stable strategy profile for conventional best reply learning, while it is easy to check that both recurrent classes are stochastically stable for M^ϵ. For conventional best reply, this result does not change in k and m – the payoff inferior equilibrium (D,r) remains the uniquely stochastically stable strategy profile. In contrast, for M^ϵ, only the payoff dominant equilibrium (U,l) is stochastically stable as soon as the sampling is sufficiently rich (consider, e.g., $m = 3$ and $k = 2$). △

6 Applications and related concepts

In this section, we conclude with some applications of dynamics with best response to noisy play and some remarks on the relation to other concepts.

Equilibria with alternative best replies As demonstrated in the previous sections, best responding to noisy play incorporates a certain degree of caution into learning behaviour that eliminates weakly dominated strategies and establishes a relation between perfect equilibria and recurrent classes or stochastically stable strategies. A particularly intuitive connection between equilibrium properties and stochastic stability can be drawn for games with alternative best replies. Consider, e.g., the following *matching market*. There are three hospitals (h_1, h_2, h_3) with strict preferences over sets of students (s_1, s_2, s_3) who wish to conduct an internship at the hospital. Students in turn have strict preferences over hospitals. To be specific, let preferences be given by the following lists.

$$h_1 : s_1, (s_2, s_3), s_2, \emptyset; \qquad s_1 : h_3, h_2, h_1$$

$$h_2 : s_2, (s_1, s_3), s_3, \varnothing; \quad s_2 : h_1, h_3, s_2$$

$$h_3 : s_3, (s_1, s_2), s_1, \varnothing; \quad s_3 : h_2, h_1, h_3.$$

That is, hospital h_1 strictly prefers student s_1 as an intern and if s_1 is not available, the hospital would prefer to have both s_2 and s_3 as interns (e.g., consider students with different specialisation) rather than only s_2. s_3, however, is unacceptable for the hospital in the absence of s_2 (e.g., because s_3's specialisation is a nice add on but other specialisations are needed for the daily business of the hospital). Now suppose that the matching between hospitals and students is retrieved by the so-called (student proposing) deferred acceptance algorithm:[11] Students submit their preferences over hospitals and hospitals submit their preferences over (sets of) students. In the first round of the algorithm, students apply at the hospital they rank highest and hospitals accept their most preferred set of applications and reject the remaining applications. Rejected students then apply at the next best hospital and hospitals choose the most preferred set of students given new applications and applications they accepted in the previous round. The algorithm terminates if no student is rejected. For example, if students and hospitals submit their actual preferences in the example, the algorithm terminates after the first round as all students apply at their top-ranked hospital and are accepted. If students submit actual preferences while at least two hospitals only declare their top-ranked student acceptable, the algorithm (after several rejections) terminates in a matching of h_1 with s_1, h_2 with s_2, and h_3 with s_3. This matching is strictly better for each hospital than the matching that results from truthful preference revelation. But if students and two hospitals reveal truthfully while one hospital declares only the top-ranked student acceptable, the corresponding hospital remains unmatched in the algorithm.

If students submit truthfully and hospitals decide between truthful preference submission and a submission of a truncated list that only declares the top-ranked student acceptable, it is therefore a strict Nash equilibrium that hospitals truthfully submit and a Nash equilibrium with alternative best replies that all hospitals submit a truncated list – whenever the other two hospitals submit a truncated list, it is irrelevant whether the third hospital truncates or truthfully submits.

For conventional best response learning as in Procedure 1, only truthful preference submission resembles a recurrent class - the alternative best response to truthfully submit while two other hospitals truncate destabilizes truncation (similar to Example 6). For best response to noisy play, however, it is a strict best response to truncate if the other two hospitals truncate (and it can be shown that only truncation is stochastically stable as it takes fewer mistakes for the dynamics to exit states with only truthful revelation than to exit states with only truncation).

[11] This design is used in the National Resident Matching Program for medical students in the U.S. (see Roth and Peranson, 1999).

A similar situation has been analysed in a laboratory study by Charness and Jackson (2007). In this experiment, the decision of a row and a column player in a stag-hunt game is formed by pairs of two individuals, and the treatment variation is the quota needed for the implementation of a strategy. In one treatment, it needs two votes for stag to be played by the team and in the other treatment one vote for stag is sufficient, hare is played otherwise. If a quota of one is needed for stag, there is the alternative best response to vote for hare if the other individual of a team votes for stag. Likewise, if a quota of two is needed for stag, there is an alternative best response to vote for hare if the other individual votes for hare. As a consequence, voting for stag is a strict equilibrium for quota two and an equilibrium with alternative best replies for quota one – and vice versa for hare. In the experiment, participants predominately choose stag if the quota is one and hare if the quota is two, i.e., they choose the equilibrium with alternative best replies rather than the strict equilibrium.

This popularity of an equilibrium strategy with alternative best response suggests that truncation (the unique stochastically stable strategy in a dynamics with best response to noisy play) is more likely than truth-telling (the unique recurrent class for conventional best response learning) and provides some empirical support for the cautious learning behaviour modelled in this paper. Charness and Jackson (2007) also introduce *robust belief equilibria* that are based on a learning dynamic with an anticipation of trembles to explain their findings. However, the selection among best replies proposed in Charness and Jackson (2007) is tailored to their experimental set-up that focuses on the selection between pure strategy equilibria and can not be straightforwardly extended to an u.h.c. best response correspondence against noisy play for arbitrary games (see also Footnote 25 in Charness and Jackson, 2007).

Generic sequential two-player games On first sight, the motivating example in Figure 1 seems a bit contrived as it is a non-generic normal form game. But there is a natural class of games with such a non-generic normal form: the class of generic extensive form games. In these games no player is indifferent between any two distinct terminal histories. Consider game Γ_7 as an illustration.

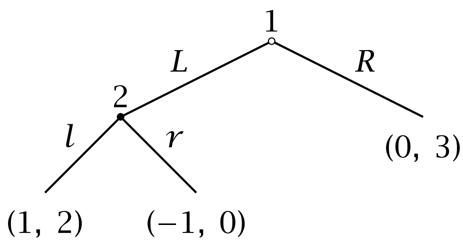

Figure 7: A two-stage game.

The extensive form in Figure 7 gives rise to the normal form depicted in Figure 8.

	l	r
L	$1,2$	$-1,0$
R	$0,3$	$0,3$

Figure 8: The associated normal form to Γ_7.

Strategy l weakly dominates r for player 2 such that the latter is never chosen in a dynamic with best response to noisy play. Only observing the play of l, player 1 best replies with L such that the backward induction outcome (L, l) resembles the unique recurrent strategy profile. This observation straightforwardly generalises to all generic extensive form games for two players with perfect information where each player only moves once. The learning dynamic with best response to noisy play (applied to the associated normal form) always has a singleton recurrent class consisting of the (unique) backwards induction outcome. Just observe that the second mover has a unique admissible strategy for every history and for generic games the first mover has a unique best response.

Dekel-Fudenberg procedure The assumption that players maximise expected utility holding beliefs that have full support in the set of other players' strategies has already been investigated by Börgers (1994). He demonstrates that a strategy is chosen under approximate common knowledge of this behaviour and beliefs if and only if the strategy survives an iterated elimination procedure introduced by Dekel and Fudenberg (1990). This procedure first eliminates all weakly dominated strategies and then iteratively eliminates strictly dominated strategies. Hence, caution (as expressed by beliefs with full support and (approximate) common-knowledge thereof) allows to include admissibility into the connection between common knowledge of rationality and iterated elimination of strictly dominated strategies as established by Bernheim (1984) and Pearce (1984) (see Brandenburger, 1992 for an overview). Recurrent classes of a dynamic with best response to noisy play only contain strategies that survive the iterated elimination procedure by Dekel and Fudenberg (1990). But as indicated by several of our examples, not all strategies that survive the iterated elimination procedure are elements of recurrent classes for a given specification of noisy play. Consider, e.g., the perfect equilibrium strategies D and r in Example 5 for uniform trembles in Procedure 2.

Semi-robust best responses An earlier attempt to integrate cautious behaviour into best response learning can be found in Hurkens (1995) who assumes that players play semi-robust best replies as introduced by Balkenborg (1992) rather than best replies as in the model by Young (1993). A semi-robust best reply s_i to a mixed strategy profile σ_{-i} is a best reply against σ_{-i} and also a best reply to an open subset of any neighbourhood of σ_{-i}. As there always exists a semi-

	h	t	g
H	$-1,1$	$0,0$	$-1,1$
T	$0,0$	$-1,1$	$-1,1$

Figure 9: A game with a unique proper equilibrium and multiple perfect equilibria.

robust best reply (see Balkenborg, 1992), this augments the learning process of Young (1993) with a well-defined selection rule in case of multiple best replies. As semi-robust strategies are admissible, this augmented learning process also guarantees the admissibility of recurrent (or stochastically stable) strategies. However, recurrent classes of the augmented learning process in Hurkens (1995) do not coincide with recurrent classes for a dynamic process with best response to noisy play. For instance, both perfect equilibria in Example 5 are recurrent in a dynamic with semi-robust best replies, while the selection in a dynamic process with best response to noisy play depends on the specification of noise as discussed in Section 4. More importantly, a semi-robust best response correspondence is not necessarily convex-valued which rules out the fixed point arguments made in Section 4 to establish a relation between trembling hand perfection and recurrent classes (see the corresponding discussion on p.170 in Balkenborg, Hofbauer, and Kuzmics, 2013).

Properness Imposing restrictions beyond Definition 1 on the noisy estimation procedure also allows to investigate the relation between recurrent classes and proper equilibria. For a normal form game and a parameter $\epsilon > 0$, a totally mixed strategy profile σ is *ϵ-proper* if for all player i, $\sigma_i(s_i)$ is at most ϵ times the probability $\sigma_i(s_i')$ whenever $u_i(s_i, \sigma_{-i}) < u_i(s_i', \sigma_{-i})$. A strategy profile σ^P is a *proper equilibrium* if it is the limit of a sequence of ϵ-proper strategy profiles (see Myerson, 1978). As a proper equilibrium is trembling hand perfect, Proposition 2 implies that for every proper equilibrium in pure strategies there is a noisy estimation procedure such that the proper equilibrium profile constitutes a singleton recurrent class. As in the case of trembling hand perfection (see Section 4) the relation is less clear-cut for equilibria in mixed strategies. The following example demonstrates how a learning process with noisy estimates based on Procedure 3 selects the (unique) proper equilibrium out of a continuum of trembling hand perfect equilibria.

Example 7. Consider the game Γ_9 depicted in Figure 9. All mixed strategy profiles where player 2 chooses g (regardless how player 1 mixes between H and T) are perfect equilibria but only the perfect equilibrium where player 1 chooses H and T with equal probability is proper. Regarding the recurrent classes of a learning dynamic with best response to noisy play, observe that h and t are weakly dominated such that only g can be a recurrent strategy. Now consider player 1 and assume that $m = 2$ and $k = 1$. In the long run, player 1 will always sample g as player 2's choice. If player 1's noisy estimate has the logit component as in Procedure 3, she will expect player 2 to tremble more often to h whenever she played H in her sample and to tremble more often to t whenever she played T in her sample. As a consequence, player 1 will choose H whenever she played T in the sample and vice versa such that both strategies will be equally likely in the long run as it is also the case in the unique proper equilibrium. △

However, the example also indicates that the selection of the proper equilibrium depends on details of the learning process. If, e.g., players estimate according to Procedure 2 but player 1 considers a tremble of player 2 to h more likely than a tremble to t, T is the unique best response to noisy play to g and a perfect equilibrium that is not proper (T, g) constitutes the unique recurrent class. However, it can be shown that there is always a noisy estimation procedure such that a given perfectly CURB set owns the support of a proper equilibrium. This establishes a *path to properness* analogously to Corollary 1 (provided that the game is weakly non-degenerate in best replies and the dynamic process has rich sampling, see Ritzberger and Weibull, 1995, who derive a similar result for a general class of continuous time learning processes).

A Appendix

Proof of Lemma 3. Consider a learning dynamic with best response to noisy play M^ϵ and a recurrent class A. Observe that for any player i any recurrent strategy s_i played in A will be arbitrarily often in a sufficiently long memory. And observe that the regular grid of frequencies $Q_{ji} = \{0, \frac{1}{k}, \ldots, \frac{k-1}{k}, 1\}$ that player j can sample regarding the occurrence of a strategy $s_i \in S_i$ contains any rational number in $[0, 1]$ if k is sufficiently large. Hence, for every $\sigma_i \in \Theta_i$ and $\delta > 0$ there is $\widetilde{k}$ and $\widetilde{m}$ such that a sampled distribution $q_{ji} \in Q_{ji}$ is in a δ-neighbourhood of σ_i in Θ_i whenever $k \geq \widetilde{k}$ and $m \geq \widetilde{m}$. As a consequence, for any $\sigma \in \Theta$ and $\delta > 0$ there is $\widetilde{k}$ and $\widetilde{m}$ such that a sampled distribution of play is within a δ-neighbourhood of σ in Θ.

(Minimal β^ϵ-CURB $\Rightarrow$ recurrent): Suppose $Y \subseteq S$ is a set of strategies which is not recurrent for arbitrary $k > \widetilde{k}$ and $m > \widetilde{m}$. Then there is $y \in Y$ such that y is transient, i.e., in any recurrent class, y will eventually cease to be played and y is not a best response to noisy play to arbitrary probability distribution over recurrent strategies that can be sampled. But for $k > \widetilde{k}$ and $m > \widetilde{m}$, there is a sampled distribution in any neighbourhood of $\sigma \in \Theta$ and as long as we consider games that are weakly non-degenerate in best replies, there is a sampled distribution with the same best response to noisy play as against σ. Hence, if y is transient, it cannot be in a *minimal* β^ϵ-CURB set.

(Recurrent $\Rightarrow$ minimal β^{ϵ}-CURB): Suppose that $Y \subseteq S$ is a set of strategy profiles which are recurrent. Recall that for $k > \widetilde{k}$ and $m > \widetilde{m}$ there is a sampled distribution in any neighbourhood of $\sigma \in \Theta$ and as long as we consider games that are weakly non-degenerate in best replies, there is such a sampled distribution with the same best response to noisy play as against σ. Then, Y is also minimally β^{ϵ}-CURB. ■

Proof of Proposition 2. As s^p is perfect there is $\left\{\sigma^{\ell}\right\}_{\ell=1}^{\infty} \subseteq \operatorname{int}(\Theta)$, with $\left\{\sigma^{\ell}\right\}_{\ell=1}^{\infty} \to s^p$ such that for each ℓ and each $i \in N$, $s_i^p \in \mathcal{B}_i\left(\sigma_{-i}^{\ell}\right)$. Now define a best response to noisy play correspondence $\beta_i^{\epsilon}(\sigma_{-i}) = \beta_i(\epsilon\sigma_{-i} + (1-\epsilon)\sigma_{-i}^{\ell})$. By construction, $s_i^p \in \mathcal{B}_i^{\epsilon}(\sigma_{-i}^p)$. If there is $\widetilde{\epsilon} > 0$ such that $\{s_i^p\} = \mathcal{B}_i^{\epsilon}(\sigma_{-i}^p)$ for all $i \in N$ and $\epsilon < \widetilde{\epsilon}$, then s^p constitutes a singleton recurrent class for sufficiently small noise and we are done. If, for some $i \in N$, there is $s_i' \in \mathcal{B}_i^{\epsilon}(\sigma_{-i}^p)$ for a neighbourhood of $\epsilon = 0$, we need to augment the best response to noisy play as follows. As we assume that s_i^p and s_i' are not own-payoff equivalent and s_i^p is a perfect equilibrium strategy and thereby admissible, there is $s_{-i}' \in S_{-i}$ such that $u_i\left(s_i^p, s_{-i}'\right) > u_i\left(s_i', s_{-i}'\right)$. Now let $\beta_i^{\epsilon}\left(\sigma_{-i}\right) = \beta_i\left(\left(1-\epsilon\right)\right)\sigma_{-i} + \epsilon\left(\left(\epsilon s_{-i}' + \left(1-\epsilon\right)\sigma_{-i}^{\ell}\right)\right)$. For ϵ sufficiently small, $s_i^p \in \mathcal{B}_i^{\epsilon}(\sigma_{-i}^p)$ but $s_i' \notin \mathcal{B}_i^{\epsilon}\left(\sigma_{-i}^p\right))$. If now $\left\{s_i^p\right\} \neq \mathcal{B}_i^{\epsilon}\left(\sigma_{-i}^p\right)$ for some $i \in N$, iterate the preceding argument. ■

Proof of Proposition 3. Consider a sequence $\left\{\epsilon^{\ell}\right\}_{\ell=1}^{\infty}$ with $\epsilon^{\ell} \to 0$ and let $\left\{Y^{\ell}\right\}_{\ell=1}^{\infty}$ be a sequence of subsets of S such that Y^{ℓ} is a minimal $\beta^{\epsilon^{\ell}}$ CURB set. For each ϵ^{ℓ} and $i \in N$, $\beta_i^{\epsilon^{\ell}}$ is a non-empty, convex-valued, and u.h.c. correspondence mapping the compact and convex set Θ_{-i} to the compact and convex set Θ_i. By definition of a β^{ϵ}-CURB set, restricting $\beta_i^{\epsilon^{\ell}}$ to Y^{ℓ} is also a non-empty, convex-valued, and u.h.c. correspondence mapping the compact and convex set Υ_{-i}^{ℓ} to the compact and convex set Υ_i^{ℓ} with $\Upsilon^{\ell} = \Delta(Y^{\ell})$. By Kakutani's theorem, there is a fixed point σ^{ℓ} of $\beta^{\epsilon^{\ell}}$ on Υ^{ℓ} for each ℓ. By definition of a β^{ϵ}-CURB set σ^{ℓ} is also a fixed point of $\beta^{\epsilon^{\ell}}$ on Θ. Hence, the sequence $\left\{Y^{\ell}\right\}_{\ell=1}^{\infty}$ induces a sequence of fixed points $\left\{\sigma^{\ell}\right\}_{\ell=1}^{\infty}$ of β^{ϵ} on Θ. As Θ is compact, we can restrict ourselves to a convergent subsequence $\{\sigma^k\}_{k=1}^{\infty}$ with $\sigma^k \to \sigma^p$ for some $\sigma^p \in \Theta$. As $\{\sigma^k\}_{k=1}^{\infty}$ converges to σ^p, we get for all $i \in N$ and $s_i' \in \operatorname{supp}\left(\sigma_i^p\right)$ that $s_i' \in \operatorname{supp}\left(\sigma_i^k\right)$ if k is sufficiently large. Hence, σ^p is a perfect equilibrium with a support in a minimal perfectly CURB set Y. Since this result holds for any sequence $\left\{Y^{\ell}\right\}_{\ell=1}^{\infty}$ of β^{ϵ^l} CURB sets, there is a perfect equilibrium with a support in any minimal perfectly CURB set. ■

For the proof of Proposition 4, we make use of an alternative definition of a trembling hand perfect equilibrium.[12]

Definition A1. A *tremble* in a game Γ is a vector $\epsilon := (\epsilon_1, \ldots, \epsilon_{|N|})$, such that, for each $i \in N$, ϵ_i is a function $S_i \to \mathbb{R}$ satisfying that:

[12] For a proof of equivalence, see, e.g., Ritzberger (2002, Ch. 6.2).

(i) For each $s_i \in S_i$, $\epsilon_i(s_i) > 0$,

(ii) $\sum_{s_i \in S_i} \epsilon_i(s_i) < 1$.

Definition A1 allows to define an ϵ-perturbation of a strategic form game.

Definition A2. Let Γ be a game with pure strategy space S and payoffs $(u_i)_{i=1}^{n}$. The *ϵ-perturbation* of Γ is the game $\Gamma(\epsilon)$ such that for each $i \in N$, $S_i(\epsilon_i) := \{\sigma_i \in \Theta_i : \forall s_i \in S_i, \sigma(s_i) \geq \epsilon(s_i)\}$, $S(\epsilon) := \prod_{i \in N} S_i(\epsilon_i)$, $\Theta(\epsilon) := \prod_{i \in N} \Delta(S_i(\epsilon_i))$, and u_i are von Neumann-Morgenstern utilities defined on $S(\epsilon)$.

Definition A3. Let Γ be a game. A strategy profile $\sigma^p \in \Theta$ is a *(trembling hand) perfect equilibrium* whenever there are two sequences

(i) $\left\{\epsilon^{\ell}\right\}_{\ell=1}^{\infty}$ with $\left\{\epsilon^{\ell}\right\}_{\ell=1}^{\infty} \to 0$,

(ii) $\left\{\sigma^{*,\ell}\right\}_{\ell=1}^{\infty}$, with $\sigma^{*,\ell} \in \Theta\left(\epsilon^{\ell}\right)$ for each ℓ, and $\left\{\sigma^{*,\ell}\right\}_{\ell=1}^{\infty} \to \sigma^p$,

such that, for each ℓ, $\sigma^{*,\ell}$ is a Nash equilibrium of $\Gamma\left(\epsilon^{\ell}\right)$. $\sigma^{*,\ell}$ is called a *constrained* Nash equilibrium.

Definition A4 (Okada, 1981)**.** In a game Γ, a Nash equilibrium $\sigma^p \in \Theta$ is *strictly perfect* whenever for any strictly positive $\left\{\epsilon^{\ell}\right\}_{\ell=1}^{\infty}$ with $\epsilon^{\ell} \to 0$, each $\Gamma\left(\epsilon^{\ell}\right)$ has a Nash equilibrium $\sigma^{*,\ell}$ with $\sigma^{*,\ell} \to \sigma^p$.

Lemma A1. *Let σ^p be the unique perfect equilibrium of some strategic form game Γ. Then, σ^p is strictly perfect.*

The proof is similar to the proof of Theorem 1 in Okada (1981) who demonstrates that a unique equilibrium is strictly perfect (see also Remark 1, ibid).

Proof of Lemma A1. Consider a sequence of trembles $\left\{\epsilon^{\ell}\right\}_{\ell=1}^{\infty}$ with $\left\{\epsilon^{\ell}\right\}_{\ell=1}^{\infty} \to 0$ and the corresponding sequence of ϵ-perturbations of Γ, $\Gamma(\epsilon^{\ell})$ with mixed strategy space $\Theta(\epsilon^{\ell})$. Observe that $\Theta(\epsilon^{\ell})$ is a compact and convex set (a subset of the $|S|$-dimensional unit simplex defined by weak inequalities). Then, Kakutani's fixed point theorem indicates that each $\Gamma\left(\epsilon^{\ell}\right)$ exhibits a constrained Nash equilibrium σ^{ℓ}. As each σ^{ℓ} is in the compact set Θ, we can pass to a convergent subsequence $\{\sigma^{k}\}_{\ell=1}^{\infty}$ with $\{\sigma^{k}\}_{\ell=1}^{\infty} \to \sigma^{p'}$ such that $\sigma^{p'}$ is a perfect equilibrium of Γ. If $\sigma^{p'} \neq \sigma^p$, we get a contradiction to the uniqueness of σ^p. ■

Lemma A2. *Let Γ be a game with a unique perfect equilibrium s^p that is in pure strategies. If for $\epsilon \to 0$, $s_i' \in \mathcal{B}^{\epsilon}\left(s_{-i}^{p}\right)$ with $s_i' \neq s_i^p$, then s_i' and s_i^p are own-payoff equivalent strategies.*

Proof of Lemma A2. Suppose that for arbitrarily small $\epsilon > 0$, $s_i' \in \mathcal{B}^{\epsilon}\left(s_{-i}^{p}\right)$ but $s_i' \neq s_i^p$ is not an equivalent strategy. Then $u_i\left(s_i', s_{-i}^{p}\right) = u_i\left(s_i^{p}, s_{-i}^{p}\right)$, and there exists $s_{-i}' \in S_{-i}$ such that $u_i(s_i', s_{-i}') > u_i\left(s_i^{p}, s_{-i}'\right)$ (to see this recall that $s_i' \in$

$\mathcal{B}^\epsilon\left(s_{-i}^p\right)$ is admissible). Consider a sequence of trembles $\left\{\epsilon^\ell\right\}_{\ell=1}^\infty$ such that $\epsilon^\ell \to 0$ for $\ell \to \infty$ and parameterize trembles such that $\frac{\epsilon^\ell(s_j')}{\epsilon^\ell(s_j)} \to \infty$ for $s_j \neq s_j'$, $j \neq i$, and $\ell \to \infty$ (consider, e.g., trembles that are polynomial in $(1/\ell)$ with $\epsilon(s_j')$ having a strictly lower leading order in $1/\ell$ than s_j). Now note that by Lemma A1 the uniqueness of s^p implies that s^p is strictly perfect. By strict perfection, the Nash equilibria of ϵ-perturbations $\Gamma(\epsilon^\ell)$ have to converge to s^p. But as s_i' outperforms s_i^p for the given tremble specification against each σ_{-i}^ℓ for ℓ sufficiently large, the Nash equilibria of ϵ-perturbations $\Gamma(\epsilon^\ell)$ do not converge to s_i^p in contradiction to the uniqueness of s^p. ■

Proof of Proposition 4.

(i) Consider a unique perfect equilibrium s^p in pure strategies. By Lemma A1, s^p is a strictly perfect equilibrium. Hence, $s_i^p \in \mathcal{B}^\epsilon\left(s_{-i}^p\right)$ for all $i \in N$. By Lemma A2, all $s_i' \in \mathcal{B}^\epsilon\left(s_{-i}^p\right)$ are equivalent to s_i^p. Now suppose $s_i' \in \mathcal{B}^\epsilon\left(s_{-i}^p\right)$ and $s_i' \neq s_i^p$.

Case (a): s_{-i}^p is also a joint best response to $\left(s_i', s_{-i}^p\right)$. Consider a sequence $\left\{\sigma^\ell\right\}_{\ell=1}^\infty \subseteq \operatorname{int}(\Theta)$ converging to s^p such that for each $j \in N$ and each ℓ, $s_j^p \in \beta_j\left(\sigma_{-j}^\ell\right)$. Such a sequence exists by the perfection of s^p (see Definition 7). Define $\hat{\sigma}_i := (1-\alpha)\, s_i^p + \alpha s_i'$, where $\alpha \in (0,1)$. Then, $\left(\hat{\sigma}_i, s_{-i}^p\right)$ is an additional perfect equilibrium—a contradiction to the uniqueness of s^p. To see this, start from the original sequence $\left\{\sigma^\ell\right\}_{\ell=1}^\infty$ and construct a new sequence $\left\{\tilde{\sigma}^\ell\right\}_{\ell=1}^\infty \subseteq \operatorname{int}(\Theta)$ as follows: For each ℓ assign probability $(1-\alpha)\sigma_i^\ell\left(s_i^p\right)$ to s_i^p, probability $\alpha\sigma_i^\ell\left(s_i^p\right)$ to s_i', and probability $\sigma_i^\ell\left(s_i'\right)$ to s_i^p. Let $\tilde{\sigma}_i^\ell\left(s_i\right) = \sigma_i^\ell\left(s_i\right)$ for $s_i \neq s_i^p, s_i'$, and $\tilde{\sigma}_j^\ell\left(s_j\right) = \sigma_j^\ell\left(s_j\right)$ for all $s_j \in S_j$, with $j \neq i$. By construction, $\tilde{\sigma}^\ell \to \left(\hat{\sigma}_i, s_{-i}^p\right)$. By the equivalence of s_i^p and s_i' and the joint best response property of s_{-i}^p, we have $\hat{\sigma}_i \in \beta_i\left(\tilde{\sigma}_{-i}^\ell\right)$ and $s_j^p \in \beta_i\left(\tilde{\sigma}_{-j}^\ell\right)$ for $j \neq i$ for each ℓ, which implies the perfection of $\left(\hat{\sigma}_i, s_{-i}^p\right)$ in contradiction to equilibrium uniqueness.

Case (b): s_{-i}^p is not a joint best response to $\left(s_i', s_{-i}^p\right)$.Then, there is $j \neq i$ such that for some $s_j' \in S_j$

$$u_j\left(s_j', s_i', s_{-(\{i\}\cup\{j\})}^p\right) > u_j\left(s_j^p, s_i', s_{-(\{i\}\cup\{j\})}^p\right). \tag{A.1}$$

For all such $j \neq i$ one of the following is true: either there is j such that Eq. (A.2) holds or there is no such j.

$$u_j\left(s_j', s_{-j}^p\right) = u_j\left(s_j^p, s_{-j}^p\right) \tag{A.2}$$

Suppose there is no such j. Then $u_j\left(s_j', s_{-j}^p\right) < u_j\left(s_j^p, s_{-j}^p\right)$ for all $j \neq i$

for which Eq. (A.1) holds. By the continuity of $U_j(\cdot)$, the intermediate value theorem implies that, for each such j, there is $\hat{\alpha}_j \in (0,1)$ (sufficiently close to 0) such that for $\hat{\sigma}_i\left(s_i^p\right) = 1 - \hat{\alpha}_j$ and $\hat{\sigma}_i\left(s_i'\right) = \hat{\alpha}_j$ we get $U_j\left(s_j^p, \hat{\sigma}_i, s^p_{-(\{i\}\cup\{j\})}\right) > U_j\left(s_j', \hat{\sigma}_i, s^p_{-(\{i\}\cup\{j\})}\right)$. Letting $\hat{\alpha} = \min_j \hat{\alpha}_j$ (where the minimum is taken over all $j \neq i$ such that Eq. (A.1) holds), s^p_{-i} is a joint best response to $\left(\hat{\sigma}_i, s^p_{-i}\right)$ with $\hat{\sigma}_i\left(s_i^p\right) = 1 - \hat{\alpha}$ and $\hat{\sigma}_i\left(s_i'\right) = 1 - \hat{\alpha}$, and we have again constructed an additional perfect equilibrium (see Case (a)).

Now suppose there is a $j \neq i$ such that Eqs. (A.1) and (A.2) hold. Note that as s^p is the unique perfect equilibrium in the game at hand, by Lemma A1 it must be strictly perfect (cf. Definition A4).

Now consider a sequence of trembles $\left\{\epsilon^\ell\right\}_{\ell=1}^\infty$ such that $\epsilon^\ell \to 0$ for $\ell \to \infty$ and parameterize trembles such that $\frac{\epsilon^\ell(s_i')}{\epsilon^\ell(s_i)} \to \infty$ for $s_i \neq s_i'$ and $\ell \to \infty$ (without further restrictions for players other than i). Hence, for $\ell \to \infty$, i trembles into s_i' infinitely more often than into any other strategy in S_i. Strict perfection of s^p then implies that the sequence of Nash equilibria of the ϵ-perturbation $\Gamma\left(\epsilon^\ell\right)$, $\left\{\sigma^{*,\ell}\right\}_{\ell=1}^\infty$, converges to s^p. But as s_j' outperforms s_j^p for the given tremble specification against each σ_{-j}^ℓ for ℓ sufficiently large, the Nash equilibria of ϵ-perturbations $\Gamma(\epsilon^\ell)$ do not converge to s^p in contradiction to the uniqueness of s^p.

(ii) Suppose their exists a recurrent class disjoint from $\omega^p = (s^p, \ldots, s^p)$. If Γ is weakly degenerate in best replies and M^ϵ has rich sampling, recurrent classes for $\epsilon \to 0$ coincide with minimal perfectly CURB sets, which by Proposition 3 contains the support of a perfect equilibrium—a contradiction to the uniqueness of s^p.

■

References

Charalambos D. Aliprantis and Kim C. Border. *Infinite dimensional analysis*. Springer, Berlin, third edition, 2006. A hitchhiker's guide.

Carlos Alós-Ferrer and Nick Netzer. The logit-response dynamics. *Games and Economic Behavior*, 68(2):413–427, 2010.

Dieter Balkenborg. *The Properties of Persistent Retracts and Related Concepts*. PhD Thesis, University of Bonn, 1992.

Dieter Balkenborg, Josef Hofbauer, and Christoph Kuzmics. Refined best reply correspondence and dynamics. *Theoretical Economics*, 8(1):165–192, 2013.

Dieter Balkenborg, Josef Hofbauer, and Christoph Kuzmics. The refined best-response correspondence in normal form games. *International Journal of Game Theory*, 44(1):165–193, 2014.

Kaushik Basu and Jörgen W. Weibull. Strategy subsets closed under rational behavior. *Economics Letters*, 36(2):141–146, 1991.

B. Douglas Bernheim. Rationalizable strategic behavior. *Econometrica*, pages 1007–1028, 1984.

Lawrence E. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5(3):387–424, 1993.

Tilman Börgers. Weak dominance and approximate common knowledge. *Journal of Economic Theory*, 64(1):265–276, 1994.

Adam Brandenburger. Lexicographic probabilities and iterated admissibility. In Partha Dasgupta, Douglas Gale, Oliver Hart, and Eric Maskin, editors, *Economic Analysis of Markets and Games*. MIT Cambridge, Cambridge, MA, 1992.

Gary Charness and Matthew O. Jackson. Group play in games and the role of consent in network formation. *Journal of Economic Theory*, 136(1):417–445, 2007.

Eddie Dekel and Drew Fudenberg. Rational behavior with payoff uncertainty. *Journal of Economic Theory*, 52(2):243–267, 1990.

Glenn Ellison. Learning, local interaction, and coordination. *Econometrica*, 61(5): 1047–1071, 1993.

Glenn Ellison. Basins of attraction, long-run stochastic stability, and the speed of step-by-step evolution. *The Review of Economic Studies*, 67(1):17–45, 2000.

M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer-Verlag, NEW YORK, 1984.

Sjaak Hurkens. Learning by forgetful players. *Games and Economic Behavior*, 11 (2):304–329, 1995. Evolutionary game theory in biology and economics.

E. Kalai and D. Samet. Persistent equilibria in strategic games. *International Journal of Game Theory*, 13(3):129–144, 1984.

Michihiro Kandori, George J. Mailath, and Rafael Rob. Learning, mutation, and long run equilibria in games. *Econometrica*, 61(1):29–56, 1993.

Christoph Kuzmics. On the elimination of dominated strategies in stochastic models of evolution with large populations. *Games and Economic Behavior*, 72(2): 452–466, 2011.

Roger Myerson and Jörgen Weibull. Tenable strategy blocks and settled equilibria. *Econometrica*, 83(3):943–976, 2015.

Roger B. Myerson. Refinements of the Nash equilibrium concept. *International Journal of Game Theory*, 7(2):73–80, 1978.

Akira Okada. On stability of perfect equilibrium points. *International Journal of Game Theory*, 10(2):67–73, 1981.

David G. Pearce. Rationalizable strategic behavior and the problem of perfection. *Econometrica*, pages 1029–1050, 1984.

Marcin Peski. Generalized risk-dominance and asymmetric dynamics. *Journal of Economic Theory*, 145(1):216–248, 2010.

Klaus Ritzberger. *Foundations of Non-Cooperative Game Theory*. Oxford University Press, Oxford, UK, 2002.

Klaus Ritzberger and Jörgen Weibull. Evolutionary selection in normal-form games. *Econometrica*, 63(6):1371–1399, 1995.

Alvin E. Roth and Elliott Peranson. The redesign of the matching market for american physicians: Some engineering aspects of economic d. *American economic review*, 89(4):748–782, 1999.

Larry Samuelson. Stochastic stability in games with alternative best replies. *Journal of Economic Theory*, 64(1):35–65, 1994.

Reinhard Selten. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4(issue 1-2):25–55, 1975.

Eric van Damme. *Stability and Perfection of Nash Equilibria*. Springer-Verlag, New York, second edition, 1991.

Fernando Vega-Redondo. *Economics and the Theory of Games*. Cambridge University Press, Cambridge, UK, 2003.

H. Peyton Young. The evolution of conventions. *Econometrica*, 61(1):57–84, 1993.

H. Peyton Young. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, Princeton, NJ, 1998.

University of Innsbruck - Working Papers in Economics and Statistics
Recent Papers can be accessed on the following webpage:

http://eeecon.uibk.ac.at/wopec/

2015-15 **Christopher Kah, Markus Walzl:** Stochastic stability in a learning dynamic with best response to noisy play

2015-14 **Matthias Siller, Christoph Hauser, Janette Walde, Gottfried Tappeiner:** Measuring regional innovation in one dimension: More lost than gained?

2015-13 **Christoph Hauser, Gottfried Tappeiner, Janette Walde:** The roots of regional trust

2015-12 **Christoph Hauser:** Effects of employee social capital on wage satisfaction, job satisfaction and organizational commitment

2015-11 **Thomas Stöckl:** Dishonest or professional behavior? Can we tell? A comment on: Cohn et al. 2014, Nature 516, 86-89, "Business culture and dishonesty in the banking industry"

2015-10 **Marjolein Fokkema, Niels Smits, Achim Zeileis, Torsten Hothorn, Henk Kelderman:** Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees

2015-09 **Martin Halla, Gerald Pruckner, Thomas Schober:** The cost-effectiveness of developmental screenings: Evidence from a nationwide programme

2015-08 **Lorenz B. Fischer, Michael Pfaffermayr:** The more the merrier? Migration and convergence among European regions

2015-07 **Silvia Angerer, Daniela Glätzle-Rützler, Philipp Lergetporer, Matthias Sutter:** Cooperation and discrimination within and across language borders: Evidence from children in a bilingual city

2015-07 **Silvia Angerer, Daniela Glätzle-Rützler, Philipp Lergetporer, Matthias Sutter:** Cooperation and discrimination within and across language borders: Evidence from children in a bilingual city

2015-06 **Martin Geiger, Wolfgang Luhan, Johann Scharler:** When do Fiscal Consolidations Lead to Consumption Booms? Lessons from a Laboratory Experiment

2015-05 **Alice Sanwald, Engelbert Theurl:** Out-of-pocket payments in the Austrian healthcare system - a distributional analysis

2015-04 **Rudolf Kerschbamer, Matthias Sutter, Uwe Dulleck:** How social preferences shape incentives in (experimental) markets for credence goods *forthcoming in Economic Journal*

2015-03 **Kenneth Harttgen, Stefan Lang, Judith Santer:** Multilevel modelling of child mortality in Africa

2015-02 **Helene Roth, Stefan Lang, Helga Wagner:** Random intercept selection in structured additive regression models

2015-01 **Alice Sanwald, Engelbert Theurl:** Out-of-pocket expenditures for pharmaceuticals: Lessons from the Austrian household budget survey

2014-32 **Esther Blanco, Tobias Haller, James M. Walker:** Externalities in appropriation: Responses to probabilistic losses

2014-31 **Michael Kirchler, Stefan Palan:** Friendliness pays off! Monetary and immaterial gifts in consumer-salesperson interactions

2014-30 **Alice Sanwald, Engelbert Theurl:** Out-of-pocket expenditures of private households for dental services - Empirical evidence from Austria

2014-29 **Stefan Borsky, Esther Blanco:** Setting one voluntary standard in a heterogeneous Europe - EMAS, corruption and stringency of environmental regulations

2014-28 **Eberhard Feess, Christian Grund, Markus Walzl, Ansgar Wohlschlegel:** Competing trade mechanisms and monotone mechanism choice

2014-27 **Esther Blanco, E. Glenn Dutcher, Tobias Haller:** To mitigate or to adapt? Collective action under asymmetries in vulnerability to losses

2014-26 **Alice Sanwald, Thomas Schober:** Follow your heart: Survival chances and costs after heart attacks - An instrumental variable approach

2014-25 **Eric Mayer, Sebastian Rüth, Johann Scharler:** Total factor productivity and the propagation of shocks; Empirical evidence and implications for the business cycle

2014-24 **Susanne Berger, Herbert Stocker, Achim Zeileis:** Innovation and institutional ownership revisited: An empirical investigation with count data models

2014-23 **James C. Cox, Rudolf Kerschbamer, Daniel Neururer:** What is trustworthiness and what drives it?

2014-22 **Michael Kirchler, Caroline Bonn, Jürgen Huber, Michael Razen:** The "Inflow-Effect" - Trader inflow and bubble formation in asset markets *forthcoming in European Economic Review*

2014-21 **Silvia Angerer, Daniela Glätzle-Rützler, Philipp Lergetporer, Matthias Sutter:** Donations, risk attitudes and time preferences: A study on altruism in primary school children *forthcoming in Journal of Economic Behavior and Organization*

2014-20 **Christian Kleiber, Achim Zeileis:** Visualizing count data regressions using rootograms

2014-19 **Matthias Siller, Christoph Hauser, Janette Walde, Gottfried Tappeiner:** The multiple facets of regional innovation

2014-18 **Carmen Arguedas, Esther Blanco:** On fraud and certification of corporate social responsibility

2014-17 **Achim Zeileis, Christoph Leitner, Kurt Hornik:** Home victory for Brazil in the 2014 FIFA World Cup

2014-16 **Andreas Exenberger, Andreas Pondorfer, Maik H. Wolters:** Estimating the impact of climate change on agricultural production: accounting for technology heterogeneity across countries

2014-15 **Alice Sanwald, Engelbert Theurl:** Atypical employment and health: A meta-analysis

2014-14 **Gary Charness, Francesco Feri, Miguel A. Meléndez-Jiménez, Matthias Sutter:** Experimental games on networks: Underpinnings of behavior and equilibrium selection *slightly revised version forthcoming in Econometrica*

2014-13 **Uwe Dulleck, Rudolf Kerschbamer, Alexander Konovalov:** Too much or too little? Price-discrimination in a market for credence goods

2014-12 **Alexander Razen, Wolgang Brunauer, Nadja Klein, Thomas Kneib, Stefan Lang, Nikolaus Umlauf:** Statistical risk analysis for real estate collateral valuation using Bayesian distributional and quantile regression

2014-11 **Dennis Dlugosch, Kristian Horn, Mei Wang:** Behavioral determinants of home bias - theory and experiment

2014-10 **Torsten Hothorn, Achim Zeileis:** partykit: A modular toolkit for recursive partytioning in R

2014-09 **Rudi Stracke, Wolfgang Höchtl, Rudolf Kerschbamer, Uwe Sunde:** Incentives and selection in promotion contests: Is it possible to kill two birds with one stone? *forthcoming in Managerial and Decision Economics*

2014-08 **Rudi Stracke, Wolfgang Höchtl, Rudolf Kerschbamer, Uwe Sunde:** Optimal prizes in dynamic elimination contests: Theory and experimental evidence *forthcoming in Journal of Economic Behavior and Organization*

2014-07 **Nikolaos Antonakakis, Max Breitenlechner, Johann Scharler:** How strongly are business cycles and financial cycles linked in the G7 countries?

2014-06 **Burkhard Raunig, Johann Scharler, Friedrich Sindermann:** Do banks lend less in uncertain times?

2014-05 **Julia Auckenthaler, Alexander Kupfer, Rupert Sendlhofer:** The impact of liquidity on inflation-linked bonds: A hypothetical indexed bonds approach *published in* *North American Journal of Economics and Finance*

2014-04 **Alice Sanwald, Engelbert Theurl:** What drives out-of pocket health expenditures of private households? - Empirical evidence from the Austrian household budget survey

2014-03 **Tanja Hörtnagl, Rudolf Kerschbamer:** How the value of information shapes the value of commitment or: Why the value of commitment does not vanish

2014-02 **Adrian Beck, Rudolf Kerschbamer, Jianying Qiu, Matthias Sutter:** Car mechanics in the lab - Investigating the behavior of real experts on experimental markets for credence goods *forthcoming in* *Journal of Economic Behavior and Organization*

2014-01 **Loukas Balafoutas, Adrian Beck, Rudolf Kerschbamer, Matthias Sutter:** The hidden costs of tax evasion - Collaborative tax evasion in markets for expert services *forthcoming in* *Journal of Public Economics*

University of Innsbruck

Working Papers in Economics and Statistics

2015-15

Christopher Kah, Markus Walzl

Stochastic stability in a learning dynamic with best response to noisy play

Abstract
We propose a learning dynamic with agents using samples of past play to estimate the distribution of other players' strategy choices and best responding to this estimate. To account for noisy play, estimated distributions over other players' strategy choices have full support in the other players' strategy sets for positive levels of noise and converge to the sampled distribution in the limit of vanishing noise. Recurrent classes of the dynamic process only contain admissible strategies and can be characterised by minimal CURB sets based on best responses to noisy play whenever the set of sampled distributions is sufficiently rich. In this case, the dynamic process will always end up in a set of strategies that contains the support of a (trembling hand) perfect equilibrium. If the perfect equilibrium is unique and in pure strategies, the equilibrium resembles the unique recurrent class of the dynamic process. We apply the dynamic process to learning in matching markets and sequential two player games with perfect information.

ISSN 1993-4378 (Print)
ISSN 1993-6885 (Online)